




# GenBank

- The GenBank sequence database is an open access
- Annotated collection of all publicly available nucleotide sequences and their protein translations
- It is maintained by the National Center for Biotechnology Information (NCBI)
- International Nucleotide Sequence Database Collaboration (INSDC).

- 
- GenBank and its collaborators receive sequences produced in laboratories
  - Throughout the world from more than 100k distinct organisms
  - GenBank is built by direct submissions from individual laboratories, as well as from bulk submissions from large-scale sequencing centers


- 
- Original sequences
  - Direct submissions
  - Assign accession number to the sequence


- 
- The submissions are then released to the public database
  - Entries are retrievable by Entrez or downloadable by FTP



submissions of

- Expressed Sequence Tag (EST),
- Sequence-tagged site (STS),
- Genome Survey Sequence (GSS),
- High-Throughput Genome Sequence (HTGS)

- 
- GenBank direct submissions group also processes complete microbial genome sequences
  - The GenBank release notes 162.0 (October 2007)

- 
- The GenBank database includes
  - additional data sets that are constructed mechanically from the main sequence data collection,
  - and therefore are excluded from this count.

Source: <https://en.wikipedia.org/wiki/GenBank>